

# Assessing the Significance of Focal Activations Using Their Spatial Extent

K.J. Friston, K.J. Worsley, R.S.J. Frackowiak, J.C. Mazziotta, and A.C. Evans

MRC Cyclotron Unit, Hammersmith Hospital, London W12 OHS, United Kingdom (K.J.F., R.S.J.F.),  
Montreal Neurological Institute, McGill University, Montreal, Quebec, H3A 2B4 Canada  
(K.J.W., A.C.E.), and Reed Neurological Research Center, UCLA School of Medicine,  
Los Angeles, California 90024, USA (J.C.M.)

---

**Abstract:** Current approaches to detecting significantly activated regions of cerebral tissue use statistical parametric maps, which are thresholded to render the probability of one or more activated regions of *one voxel, or larger*, suitably small (e.g., 0.05). We present an approximate analysis giving the probability that one or more activated regions of *a specified volume, or larger*, could have occurred by chance. These results mean that detecting significant activations no longer depends on a fixed (and high) threshold, but can be effected at any (lower) threshold, in terms of the spatial extent of the activated region. The substantial improvement in sensitivity that ensues is illustrated using a power analysis and a simulated phantom activation study. © 1994 Wiley-Liss, Inc.

**Key words:** statistical parametric mapping, gaussian fields, activation, thresholds, excursion set, functional imaging

---

## INTRODUCTION

Functional images of the brain are almost universally compared using some form of statistical parametric mapping. Statistical parametric maps (SPMs) have voxel values that, under the null hypothesis, are distributed according to some known probability density function [Friston et al., 1990]. The most commonly used statistics are the Student's *t* [Friston et al., 1990; Worsley et al., 1992] and the correlation coefficient [e.g., Friston et al., 1993].

Any continuous probability density function (PDF) can be transformed to the Gaussian distribution or *z*-statistic. If the degrees of freedom of the original distribution are reasonably high, the resulting SPM

approximates a Gaussian field or  $SPM\{z\}$ . The analyses presented in this paper are therefore restricted to the  $SPM\{z\}$ . Some exact results for  $SPM\{t\}$ ,  $SPM\{F\}$ , and  $SPM\{\chi^2\}$  have now been derived [Worsley, in press]. The most common characterization of SPMs involves identifying significant activation foci. This is achieved by thresholding. The problem of how to correct for the multiplicity of *nonindependent* tests implicit in this approach has, in past years, been solved. The solution identifies a threshold such that for an SPM of given size, the probability of obtaining one or more *activation foci of at least one voxel*, by chance, is suitably small (for example, 0.05). This approach uses the theory of level crossings in stochastic processes [Friston et al., 1991], or the Euler characteristic [Worsley et al., 1992], applied to Gaussian processes with a known (or measurable) autocorrelation.

The threshold identified, using current techniques, takes no account of the spatial extent of an activation.

---

Received for publication September 8, 1993; accepted January 31, 1994.

Address reprint requests to K.J. Friston, MRC Cyclotron Unit, Hammersmith Hospital, DuCane Road, London W12 OHS, UK.

Activation foci are characterized not only by the threshold they reach but also by their spatial topography, for example, their shape, spatial relationships to each other, and their size. This paper deals only with size or volume. A spatially limited focus is usually considered less "significant" than a very large one [see Poline and Mazoyer, 1993 and Roland et al., 1993 for empirical studies in this area]. Clearly the detectability of significant foci would be enhanced if the volume of activated tissue were explicitly included when testing the null hypothesis that the activated region could have occurred by chance. More precisely, for any region, one would like to ask: "Over an SPM of given size, what is the probability of obtaining one or more activation foci of the same size, or larger, than the one in question?" Another way of looking at this formulation is to compare current approaches, which provide a critical value for the maximum height of a peak, to the approach proposed here, which asks: "What is the critical value for the maximum size of a focus?" The purpose of this article is to present an approximate analysis of the probability theory that is needed to answer these questions. One important aspect of assessing significance in terms of spatial extent is that the analysis is freed from the arbitrary nature of fixed and high thresholds (in the sense that significance can be assessed at any threshold).

Below we provide a theory section that defines the problem addressed, the solutions obtained, and the implications of detecting changes using a power analysis. We then describe a simple simulated phantom experiment that demonstrates the application of the theory to the detection of significant activations. More general applications of the results to real data will be deferred for later publications.

### THEORY

From a statistical perspective, three things about an SPM are of interest:

- The number ( $N$ ) of voxels above a threshold (the number of voxels in the excursion set that have values greater than a threshold  $u$ )
- The number ( $m$ ) of activated regions (clusters or connected subsets of the excursion set)
- The number ( $n$ ) of voxels in each of these clusters.

Each of these numbers has its own probability density function;  $P(N = x)$ ,  $P(m = x)$ , and  $P(n = x)$  [in this context  $P(n = x)$  is strictly a conditional probability, given that the region exists  $P(n = x | m \geq 1)$ ]. These probability functions provide a fairly complete characterization of the SPM and allow one to address a number of hypotheses.

In what follows, we apply results from the theory of Gaussian fields to a  $D$ -dimensional lattice of continuous random variables (a voxelated SPM) by considering it as a good lattice representation of an underlying continuous Gaussian field. Although the parameters  $N$  and  $n$  are, in reality, discrete variable (numbers of voxels), they are treated here like continuous measures of volume. The volumes in question are the spatial volume occupied by voxels above a certain threshold (i.e., the number of voxels in a suprathreshold cluster).

The particular probability we are interested in is: *The probability of obtaining at least one activation with  $k$  voxels or more.* This is the same as the *probability that the largest region has  $k$  voxels or more* =  $P(n_{\max} \geq k)$ , where  $n_{\max}$  is the number of voxels in the biggest region. This probability is a more general case of that currently used, namely: The probability of getting at least one activation (with one voxel or more) =  $P(n_{\max} \geq 1)$ . This is simply equal to the probability of getting at least one region  $P(m \geq 1)$ .  $P(m \geq 1)$  is usually set to 0.05 by choosing an appropriate threshold or critical height ( $u$ ). This threshold is chosen by estimating the expectation or mean of  $m(E\{m\})$  and using the fact that

$$P(m \geq 1) \leq E\{m\} \quad (1)$$

tends to equality at high thresholds [Hasofer, 1978]. An approximate expression for  $E\{m\}$  was presented in Friston et al. [1991] using a somewhat heuristic argument. More exact results will be found in Hasofer [1978], which pertain to the number of maxima. Because at high threshold ( $u$ ), the number of maxima ( $m_{\maxima}$ ) and  $m$  converge, these expressions should be equivalent (more generally  $m \leq m_{\maxima}$ ). Subsequently, the Euler characteristic ( $\chi$ ) has been used as an estimate of  $E\{m\}$  [Worsley et al., 1992]. Again, the Euler characteristic and  $m$  converge at high  $u$  (more generally,  $\chi \leq m$ ). The Euler characteristic is more amenable to mathematical analysis than the number of maxima.

The problem addressed in this article is how to estimate  $P(n_{\max} \geq k)$  and obtain a critical threshold for volume. To do this, one needs to know both the probability of obtaining an arbitrary number of regions  $P(m = x)$  and the probability that these regions have less than  $k$  voxels  $P(n < k)$ . Unfortunately, in the context of Gaussian processes, no exact results for these probabilities exist. However, using a less formal analysis, one can estimate the probability density functions using what is already known.

In the next three sections, we introduce expressions for the expectations and probability density functions

of  $m$  and  $n$  and combine these results to give the final expression for  $P(n_{\max} \geq k)$ . In brief, we use a Poisson form for  $P(m = x)$  and a standard approximation for its expectation [Hasofer, 1978; Adler and Hasofer, 1981]. The PDF for  $P(n = x)$  uses an earlier observation [Nosko, 1969a] that  $n^{2/D}$  has an exponential distribution, in conjunction with the above result for  $E\{m\}$ .

### Expectations of $N$ , $m$ , and $n$

The three variables  $N$ ,  $m$ , and  $n$  have expectations that are related:

$$E\{N\} = E\{m\} \cdot E\{n\}. \quad (2)$$

The expectation of  $N$  is known exactly because of the Gaussian univariate assumptions and is simply the appropriate integral under the normal distribution or error function  $\Phi(\cdot)$  [Friston et al., 1990]. As mentioned above, we already know the approximate expectation of  $m$ . For any threshold  $u$ , the expectations for a  $D$ -dimensional process of volume  $S$  are given by:

$$E\{N\} = S \cdot \Phi(-u) = S \int_u^\infty (2\pi)^{-1/2} e^{-z^2/2} dz \quad (3)$$

$$E\{m\} \approx S(2\pi)^{-(D+1)/2} W^{-D} u^{D-1} e^{-u^2/2} \quad (4)^1$$

$$E\{n\} = E\{N\}/E\{m\}. \quad (5)$$

$W$  is a measure of smoothness and is related to the full-width at half-maximum (FWHM) of the SPM's "resolution." Equivalently,  $W$  is inversely related to the number of "resolution elements" or Resels ( $R$ ) that fit into the total volume ( $S$ ) of the  $D$ -dimensional SPM. ( $R = S/\text{FWHM}^D$ ). More exactly, let  $\text{SPM}_i$  be the voxel value of the SPM as a function of the  $i$ th coordinate, and similarly let  $\rho_i$  be the spatial autocorrelation of the SPM as a function of the  $i$ th coordinate,  $i = 1 \dots D$ . Then:

$$W = \prod_{i=1}^D \text{Var}\{\text{SPM}_i'\}^{-1/(2D)} = \prod_{i=1}^D [-\rho_i''(0)]^{-1/(2D)} \quad (6)$$

where ' denotes the derivative [Friston et al., 1991; Worsley et al., 1992]. If the point spread function (convolved with any preprocessing filters) has a Gaussian shape with  $\text{FWHM}_1 \dots \text{FWHM}_D$  in each of the  $D$

coordinate directions, then it can be shown [Friston et al., 1991; Worsley et al., 1992] that  $W = \text{FWHM}/\sqrt{(4\log_e 2)}$  where:

$$\text{FWHM} = \prod_{i=1}^D \text{FWHM}_i^{1/D}. \quad (7)$$

In practice,  $W$  can be determined directly from the effective FWHM (if it is known) or estimated post hoc using the measured variance of the SPM's first partial derivatives according to Equation 6 [see Friston et al., 1991 and Worsley et al., 1992 for discussion and validation of this characterization of smoothness].

### Approximate expressions for $P(m = x)$ and $P(n = x)$

$P(m = x)$  has been shown, in the limit of high thresholds, to have a Poisson distribution [Adler and Hasofer, 1981, Theorem 6.9.3, p. 161]. A Poisson distribution is intuitively sensible, in the sense one can regard the centers or maxima of activated regions as multidimensional point processes with "no memory." In other words, when passing through the SPM, maxima are found in much the same way as radioactive decay events occur in time:

$$P(m = x) \approx 1/x! E\{m\}^x e^{-E\{m\}}. \quad (8)$$

Asymptotic results for the distribution and expectation of  $n^{2/D}$  have been given by Nosko [1969a,b; 1970] and are reported by Adler and Hasofer [1981, p. 158]. They show that for large thresholds,  $n^{2/D}$  has an exponential distribution with the expectation:

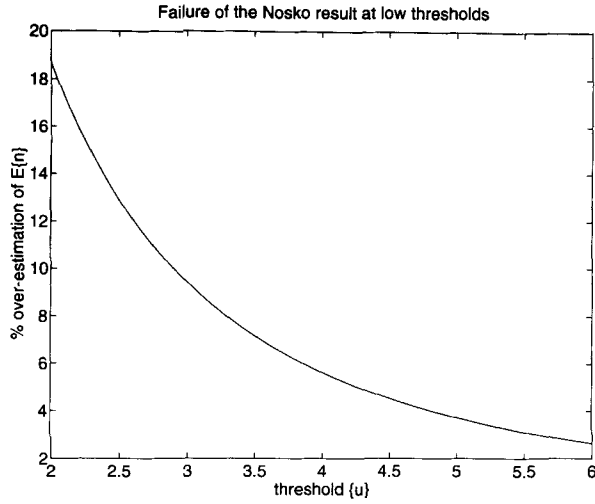
$$E\{n^{2/D}\} \approx \frac{2\pi W^2}{u^2 \Gamma(D/2 + 1)^{2/D}} \quad (9)$$

or, equivalently:

$$E\{n\} \approx \frac{(2\pi)^{D/2} W^D}{u^D}. \quad (10)$$

A simple way to derive these equations is provided at the end of this section, for the interested reader. Unfortunately, Equation 10 is not robust at lower thresholds and substantially overestimates  $E\{n\}$  when compared with estimates based on Equation 5, or indeed empirical simulations (see below). Figure 1 illustrates this point by plotting the percentage overestimation against threshold ( $u$ ). This is important be-

<sup>1</sup>[Hasofer, 1978.]



**Figure 1.**

Graphical illustration of how the Nosko result for the  $E\{n\}$ —the expectation of  $n$  (spatial extent of a region in voxels)—fails at low thresholds. The overestimation is relative to  $E\{n\}$  as defined in Equation 5. The relationship presented here does not depend on dimensionality, search volume, or smoothness.

cause we want to apply the approximations at relatively low thresholds. To obtain a better approximation for  $P(n = x)$ , we use the following device: Assume a form for  $P(n = x)$  that is asymptotically correct, and determine the parameters of the distribution by reference to its known moments. The form of  $P(n = x)$  we assume is:

$$P(n = x) \approx \frac{2\beta}{D} x^{2/D-1} e^{-\beta x^{2/D}}$$

or:

$$P(n \geq x) \approx e^{-\beta x^{2/D}}. \quad (11)$$

By a simple change of variables, it is easy to show that  $n^{2/D}$  is exponentially distributed with an expectation  $E\{n^{2/D}\} = 1/\beta$ . Similarly, the expectation of  $n$ ,  $E\{n\} = \Gamma(D/2 + 1) \cdot \beta^{-D/2}$ .  $\beta$  is determined according to the expression for  $E\{n\}$  in Equation 5:

$$\beta = [\Gamma(D/2 + 1) \cdot E\{m\}/E\{N\}]^{2/D} \quad (12)$$

In summary, we have approximate expressions for both  $P(m = x)$  and  $P(n = x)$ . These approximations allow us to estimate  $P(n_{\max} \geq k)$ .

In fact, the difference between Nosko's result and the conjecture adopted here is relatively simple (but

not necessarily small—see Figure 1) and is resolved by using a standard approximation for the error function  $\Phi(\cdot)$  that works for (and only for) high thresholds:

$$\Phi(-u) \approx \frac{e^{-u^2/2}}{u\sqrt{2\pi}}. \quad (13)$$

Substituting this into Equation 3, the expression for  $E\{N\}$  and expanding Equation 5,  $E\{n\} = E\{N\}/E\{m\}$  gives the limiting value of  $E\{n\}$ , which is given by Equation 10. Recalling that  $E\{n^{2/D}\} = 1/\beta$  (Equation 11), the same substitution can be used to verify Equation 9 using the expression for  $\beta$  in Equation 12.

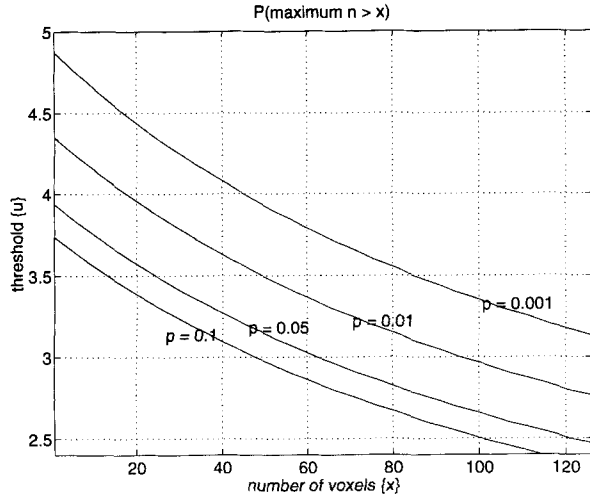
### Estimating $P(n_{\max} \geq k)$

To calculate  $P(n_{\max} \geq k)$ , we simply compute one minus the probability that all  $m$  regions have less than  $k$  voxels, times the probability of getting  $m$  regions. These probabilities are summed over all possible values of  $m$ :

$$\begin{aligned} P(n_{\max} \geq k) &= \sum_{i=1}^{\infty} P(m = i) \cdot [1 - P(n < k)]^i \\ &= 1 - e^{-E\{m\} \cdot P(n \geq k)} \\ &= 1 - \exp(-E\{m\} \cdot e^{-\beta k^{2/D}}). \end{aligned} \quad (14)$$

In the limiting case of small  $E\{m\}$ , or high threshold,  $P(n_{\max} \geq 1) \approx E\{m\}$  [note  $P(n \geq 1) = 1$ ]. This special case is precisely the one adopted in current approaches [Friston et al., 1991; Worsley et al., 1992].

Equation 14 gives an estimate of the probability of finding at least one region with  $k$  or more voxels in an SPM. For any fixed value of  $P(n_{\max} \geq k)$ , a whole family of activation foci have an equally improbable chance of occurrence. The interaction between threshold ( $u$ ) and volume ( $k$ ) is demonstrated in Figure 2, which plots the isoprobability contours of  $P(n_{\max} \geq k)$  for a two-dimensional SPM with  $128^2$  pixels (two-dimensional voxels) and FWHM = 9.42. A region of one or more pixels at a threshold of  $\sim 3.9$  is as "improbable" ( $P = 0.05$ ) as a focus with 60 or more pixels at a threshold of  $\sim 3$ . This is very important from the perspective of detecting regions of cerebral activation: Although a spatially extensive region of activation may not necessarily reach a high threshold (e.g., 3.9), it may be significant (at  $P = 0.05$ ) if assessed at a lower threshold (e.g., 3). This is the application of the above results we pursue in this article.



**Figure 2.**

Isoprobability contours calculated according to Equation 14, which give the relationship between threshold ( $u$ ) and the number of voxels ( $x$ ) an activation focus should contain to maintain a certain level of “probability.” More strictly,  $P(n_{\max} \geq k)$  (the chance probability of obtaining one or more regions of at least  $k$  voxels) has been calculated over a range of voxels and thresholds and contoured at four levels (0.1, 0.05, 0.01, 0.001). In this example,  $S$  (volume) =  $128^2$ , FWHM (smoothness) = 9.42, and  $D$  (dimensionality) = 2.

### Choice of threshold—a power analysis

The question we address here is how to choose threshold ( $u$ ). The optimum threshold should maximize sensitivity. In this section, we derive an approximate expression for the sensitivity to a “random” signal. We then demonstrate how the sensitivity, or power, depends on an interplay between the shape of the signal and the threshold used. Roughly speaking, our results will show that broader signals are best detected by low thresholds and that sharp focal signals are best detected by high thresholds.

We now give details. Suppose the “signal” resembles Gaussian kernels or foci, of random height, distributed continuously throughout the SPM. The shape of the signal is characterized by the width ( $f$ ) of these foci expressed in units of  $W$ . This signal can be modeled by a continuous ensemble of kernels with randomly distributed heights or, equivalently, by convolving an uncorrelated random process with a “kernel” of the same height. The resulting signal will be a Gaussian process of smoothness  $f \cdot W$ . Although this form of signal was chosen for theoretical convenience, it is not an unreasonable model of distributed physiological signals in the brain. Following convolution with the point spread function (and any prepro-

cessing filters), the resolution of the signal will be  $W \cdot \sqrt{(1 + f^2)}$ . Let the convolved signal have a standard deviation  $\sigma$ , where  $\sigma$  corresponds to the amplitude of the measured signal. Note that  $\sigma$  characterizes the amount of *measured* signal, not the underlying physiological signal (which would be subject to partial volume effects if  $f$  were small).

Power is the probability of getting at least one true-positive, while using criteria that protect against false-positives. The power at a particular threshold can be estimated with the probability of detecting at least one “signal” using a critical region size ( $k_\alpha$ ) such that  $P(n_{\max} \geq k) = \alpha$ , where  $\alpha$  is suitably small, say 0.05.  $k_\alpha$  is found by inverting Equation 14:

$$k_\alpha \approx [\log(-E\{m\}/\log(1 - \alpha))/\beta]^{D/2}. \quad (15)$$

A small table of  $k_\alpha$  is provided as a reference for anyone wishing to duplicate our calculations (Table I). This size criterion  $k_\alpha$ , at threshold  $u$ , is now applied to the process representing noise plus signal. This process will have zero mean and variance  $1 + \sigma^2$ , which means the original threshold  $u$  is effectively:

$$u^* = u/\sqrt{1 + \sigma^2}. \quad (16)$$

The process with signal will itself be a smooth Gaussian field with an effective resolution ( $W^*$ ),

$$W^* = W \cdot \sqrt{\frac{1 + \sigma^2}{1 + \sigma^2/(1 + f^2)}} \quad (17)$$

The validity of Equation 17 is easily established by deriving the autocorrelation function of the signal plus noise, in terms of  $W$ , and using Equation 6. The expectations of  $N$  and  $n$  for signal plus noise ( $E\{N\}^*$  and  $E\{m\}^*$ ) are given by substituting  $W^*$  and  $u^*$  in Equations 3 and 4.  $\beta^*$  is similarly derived according to Equation 12. The probability of getting at least one region bigger than  $k_\alpha$  in the new process is given by Equation 14. Because this region is unlikely ( $P \leq \alpha$ ) to be due to the noise component, one can interpret the result as the probability of getting at least one true-positive. Therefore, on substituting the above:

$$\text{power} = 1 - \exp[-E\{m\}^* \cdot \exp(-\beta^* k_\alpha^{2/D})]. \quad (18)$$

The results of this kind of analysis are presented graphically in Figure 3, which shows how power depends on threshold and region size for a three-dimensional search volume of 65,536 voxels with FWHM = 6.12,  $\sigma = 0.7$ , and a false-positive rate  $\alpha =$

TABLE I. Critical sizes of  $k^*$

Threshold ( $u$ )	FWMH = 9.4			FWHM = 9.2			FWHM = 6.2		
	$S = 4,096, D = 1$			$S = 16,384, D = 2$			$S = 65,536, D = 3$		
	Alpha								
	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
2.4	12	13	15	110	131	179	283	476	711
2.6	10	11	13	85	103	144	272	344	527
2.8	9	10	12	64	80	116	191	247	390
3.0	8	9	11	46	60	92	131	174	287
3.2	6	7	9	31	44	72	86	120	209
3.4	5	6	8	19	30	55	53	79	149
3.6	3	5	7	8	18	41	30	49	104
3.8		3	6		8	28	13	27	69
4.0			4			17	3	12	43
4.2			3			8		3	24
4.4									11
4.6									2

\*Approximate critical values  $k_c$  of  $k$ , the number of voxels in a suprathreshold cluster, chosen such that  $P(n_{\max} \geq k) \approx \alpha$ , where  $n_{\max}$  is the number of voxels in the largest cluster. These values are specific to the tabulated threshold ( $u$ ), resolution (FWHM), search volume ( $S$ ), and dimensionality ( $D$ ). They are provided as a reference for those who wish to reproduce our calculations.

0.05. For  $f < 0.7$  (the underlying signals are narrower than 70 percent of the resolution of the SPM), then power increases with threshold, so that the most powerful test is just that based on the maximum value at a voxel, as advocated by Friston et al., [1991] and Worsley et al., [1992]. If  $f > 0.7$  (the activated region is broader than 70 percent of the resolution), then power increases as the threshold is lowered. The results presented here are only good approximations for large thresholds, so it seems prudent to keep the thresholds high enough for the  $P$  values to remain valid. Our simulations (see below) suggest that the approximations hold reasonably well for thresholds as low as 2.4.

### Empirical verification of the PDFs

This section presents an empirical validation of the above results. This is necessary because the expressions introduced depend on a number of approximations and should be regarded as (mathematically) rough estimates, especially at lower thresholds.

We present one-, two-, and three-dimensional simulations to assess the adequacy of Equations 8 and 11 to describe  $P(m = x)$  and  $P(n = x)$  and to validate Equation 14, the expression for  $P(n_{\max} \geq k)$ . The one-dimensional simulations used processes with 4,096 voxels

( $S = 4,096$ ), a smoothness (FWHM) of 9.4, and a threshold ( $u$ ) of 2.6. The two-dimensional simulations used processes with  $256 \times 256$  pixels, FWHM = 9.2, and  $u = 2.58$ . The parameters for the  $32 \times 32 \times 64$  voxel, three-dimensional simulated processes were FWHM = 5.7 and  $u = 2.8$ . For each of the three simulations,  $10^4$

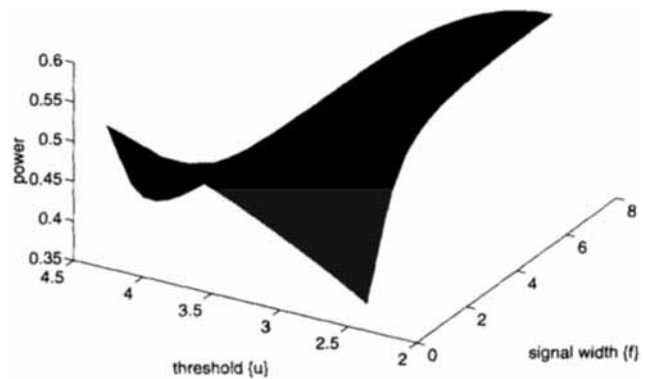


Figure 3.

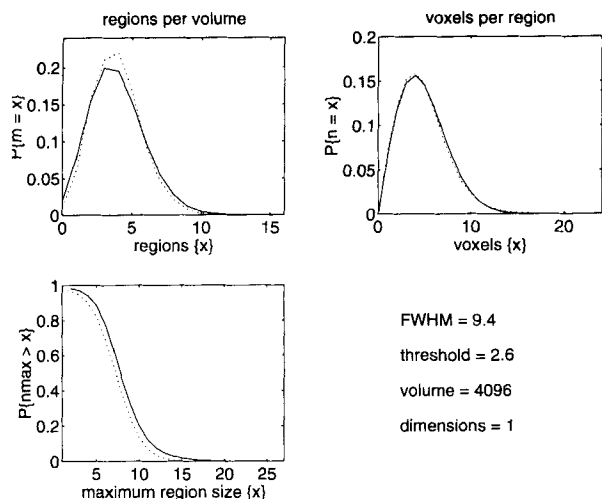
Two-dimensional plot of the power as a function of threshold ( $u$ ) and the width of a signal ( $f$ ) expressed in units of the SPM's smoothness ( $W$ ). The key thing to note here is the marked dissociation between the optimum thresholds (thresholds that maximize power) at high and low signal widths ( $f$ ).

realizations were created using uncorrelated Gaussian processes, which were convolved with (one-, two-, or three-dimensional) Gaussian kernels. The FWHMs given above are based on post hoc empirical estimates of  $W$  using Equation 6. For each realization, the number of suprathreshold regions ( $m$ ), the numbers of voxels in each region ( $n$ ), and the size of the largest region ( $n_{\max}$ ) were recorded. Using this data, we compared the empirical and theoretical estimates of  $P(m = x)$ ,  $P(n = x)$ , and  $P(n_{\max} \geq k)$ .

The agreement between the empirical (dotted lines) and theoretical (solid lines) distributions is evident (Figure 4: one-dimensional, Figure 5: two-dimensional, and Figure 6: three-dimensional). In all simulations, the agreement is particularly good for large clusters of voxels. The values of  $P(n_{\max} \geq k)$  are too conservative in the 1-dimensional case. However, at higher dimensions they cannot be distinguished from the empirically determined values.

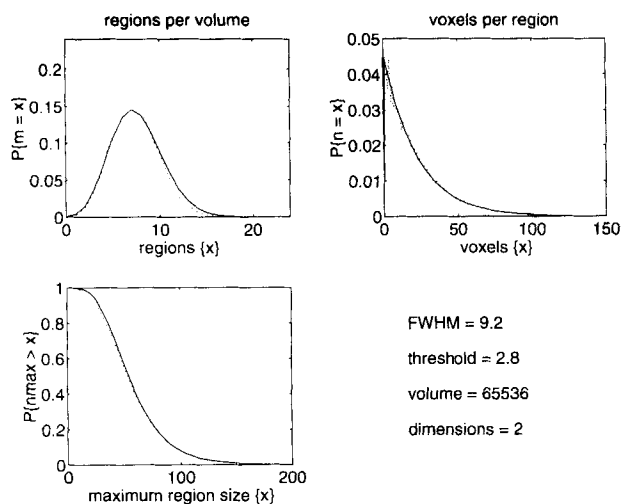
The small discrepancies between the empirical and theoretical distributions may derive from a number of sources:

- Many of our approximations were only asymptotically true, in the limit of high thresholds.



**Figure 4.**

Empirically (dotted lines) and theoretically (solid lines) determined probability density functions for the number of regions ( $m$ ) per realization (top left); the number of voxels per region ( $n$ ) (top right); and  $P(n_{\max} \geq k)$  the probability that the largest region, per realization, has  $k$  or more voxels (bottom left). In this one-dimensional simulation, FWHM (smoothness) = 9.4,  $u$  (threshold) = 2.8, and  $S$  (volume) = 4096. The empirical estimates are from 10,000 realizations. The most marked failure in these one-dimensional simulations was an overestimation of the probability of getting a large region by chance. This failure is in the conservative direction.



**Figure 5.**

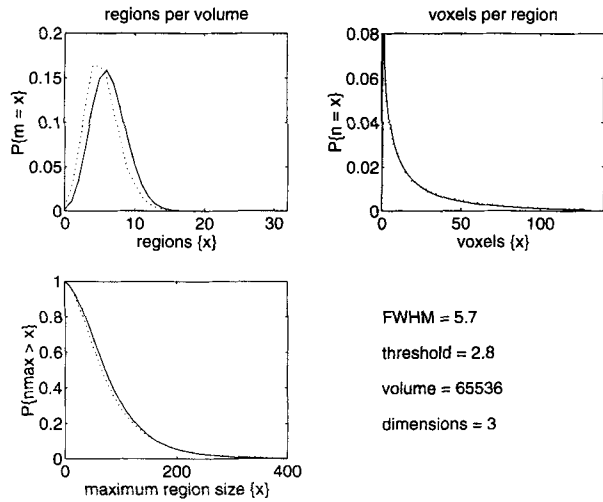
As for Figure 2, but using two-dimensional processes. FWHM (smoothness) = 9.2,  $u$  (threshold) = 2.58, and  $S$  (volume) = 256 × 256. The empirical estimates are from 10,000 realizations. The agreement between the empirical and theoretical values for  $P(n_{\max} \geq k)$  is remarkably good, particularly at large regions sizes (empirical, dotted lines; theoretical, solid lines).

- Using discretized processes (voxels) had an effect.
- Our simulated process was not “very large” and had edges.

### SIMULATED PHANTOM EXPERIMENTS

To provide a concrete and clear illustration of how these equations can be applied to detecting activation foci, we performed a simulated phantom activation study. We compared two images from a phantom, 108 pixels in diameter, with six circular wells. In the *baseline* condition, the wells contained the same activity as in the main body of the phantom. This background activity was 100 counts/pixel. The counts/pixel in a typical human study are about 6 to 12; therefore, the simulated images can be thought of as coming from about ten subjects. In the *activation* phantom, the wells contained activity that was 10 percent higher than the background. The wells increased in size with diameters of  $W \cdot (2 + j) / \sqrt{2}$ , where  $j$  ran from one to six. The smallest well was therefore about the size of the FWHM and subject to partial volume effects.

Matrix manipulations and computations were performed in Matlab (MathWorks, Inc., Sherborn, MA). The 128 × 128 voxel baseline and activated images were constructed assuming Poisson counting statistics (variance equal to mean activity) and a Gaussian point



**Figure 6.**

As for Figure 2, but using three-dimensional processes. FWHM (smoothness) = 5.7,  $u$  (threshold) = 2.8, and  $S$  (volume) =  $32 \times 32 \times 64$ . The empirical estimates are from 10,000 realizations. As for the two-dimensional case, the agreement is remarkable (empirical, dotted lines; theoretical, solid lines).

spread function with  $\sigma = 3$ . The simulated resolution corresponded to an FWHM of 7.06 pixels. Figure 7 (top) shows the two images. The baseline and activated images were subtracted and scaled to unit variance. Figure 7 (bottom) shows the subtraction image and the underlying signal (signal from the wells following convolution). The significance of the six activation foci were assessed using two methods:

- A threshold was applied to the normalized difference image according to current approaches, which rendered  $P(m \geq 1) = P(n_{\max} \geq 1) \approx 0.05$ . In this instance, the threshold was 4.108. This analysis only detected the three largest activation foci. The results of this conventional and high thresholding are shown in Figure 8 (left).
- The normalized difference image was thresholded at a much lower and arbitrary level (2.8). The significance of each focus was assessed with the  $P(n_{\max} \geq k)$  calculated using Equation 14. The results of this analysis are shown in Figure 8 (right). Although detected, the smallest focus is not significant [ $P(n_{\max} \geq k) = 0.75$ ]. Indeed, something this size or larger would occur at least once on about 75 percent of occasions by chance. The second smallest focus had a  $P(n_{\max} \geq k)$  of 0.17. The remaining four foci were found to be highly significant.

These results demonstrate a substantial increase in sensitivity to actual change when the activation topog-

raphy is taken into account. This approach frees the analysis of significant focal change from the arbitrary nature of high thresholds.

The configuration of activation foci we chose to report here is rather arbitrary. We obtained equivalent results with simulated activation foci of constant shape but varying intensity and using combined differences in size and intensity.

## DISCUSSION

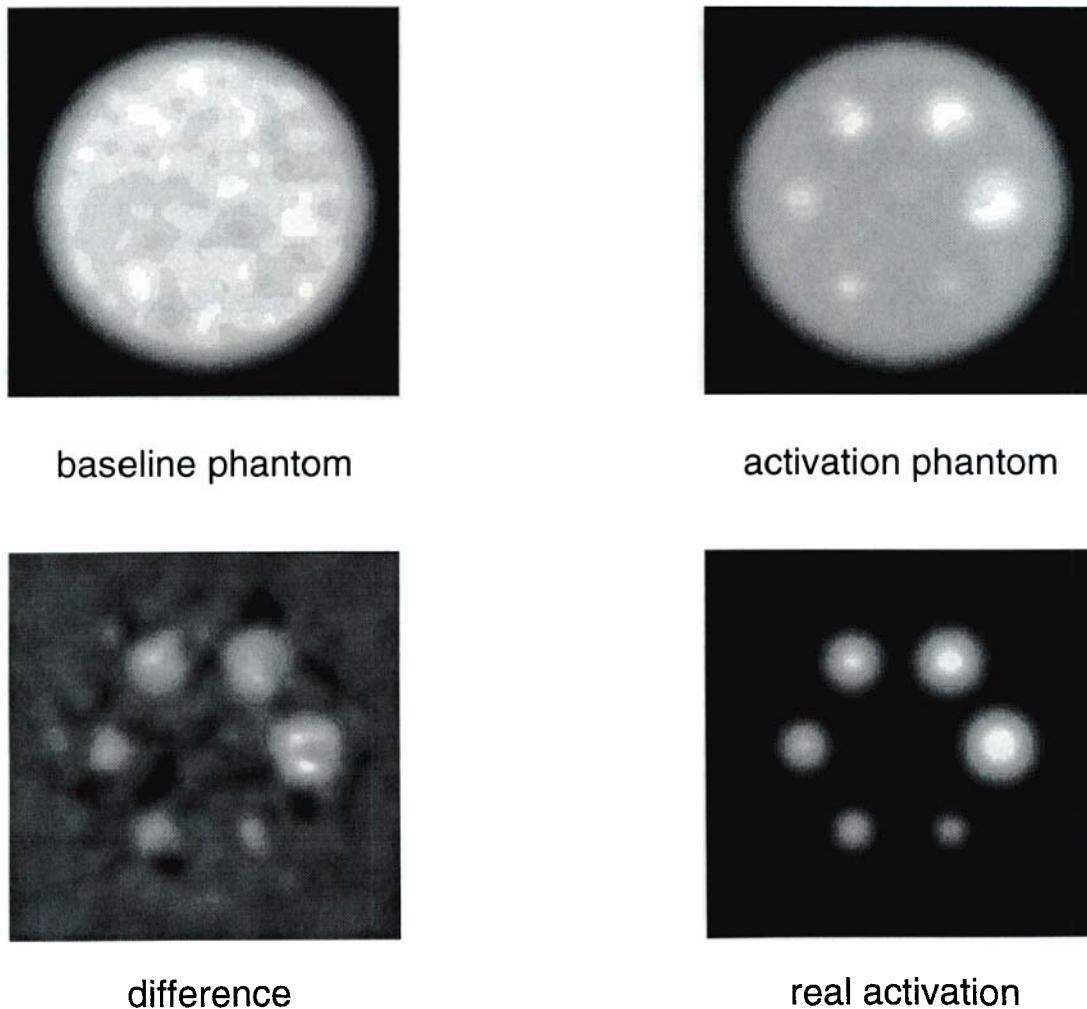
We have used approximate expressions for the probability density functions of the number of regions above threshold, and the number of voxels in each region. This characterization of Gaussian processes is rough but fairly complete and allows a significant advance in detecting activation foci in SPMs. Previous approaches have taken an activation focus, at some threshold, and estimated the probability that one or more such foci, of any size, could have occurred by chance. The threshold is usually set such that this probability is 0.05. We are now in a position to estimate the significance of an activation focus in terms of the probability that one or more foci, of the *same or greater size*, could have occurred by chance (by using Equation 14). This is important because information about the spatial extent or volume of the activation is explicitly included, and the analysis of significant focal change is freed from the arbitrary nature of high thresholds. The improved sensitivity that results allows the threshold to be lowered to much more realistic levels with no increase in the experiment-wise probability of a false-positive.

### Thresholds and signal width

A power analysis based on the results presented here suggests that narrow focal activations are most powerfully detected by high thresholds (as currently implemented), whereas broader, more diffuse activations are best detected by low thresholds. Alternately, for a fixed threshold, *power increases with resolution*. The fundamental importance of this for functional magnetic resonance imaging (MRI) studies is obvious; however, it also raises issues about the optimum resolution for positron emission tomography (PET) data, for which the current practice is to use low resolutions.

The proposal to use lower thresholds is vindicated by the power analysis, but is predicated on the assumption that real signals are broader than the resolution. Empirical evidence suggests that this is the case: The autocorrelation of physiological changes





**Figure 7.**

Simulated  $128^2$  phantom data. **Top left.** Baseline phantom image 108 pixels in diameter with uniform activity of about 100 counts/pixel and an FWHM of 7.06 pixels. **Top right.** Activated image in which six wells, of increasing size, have been filled with activity 10

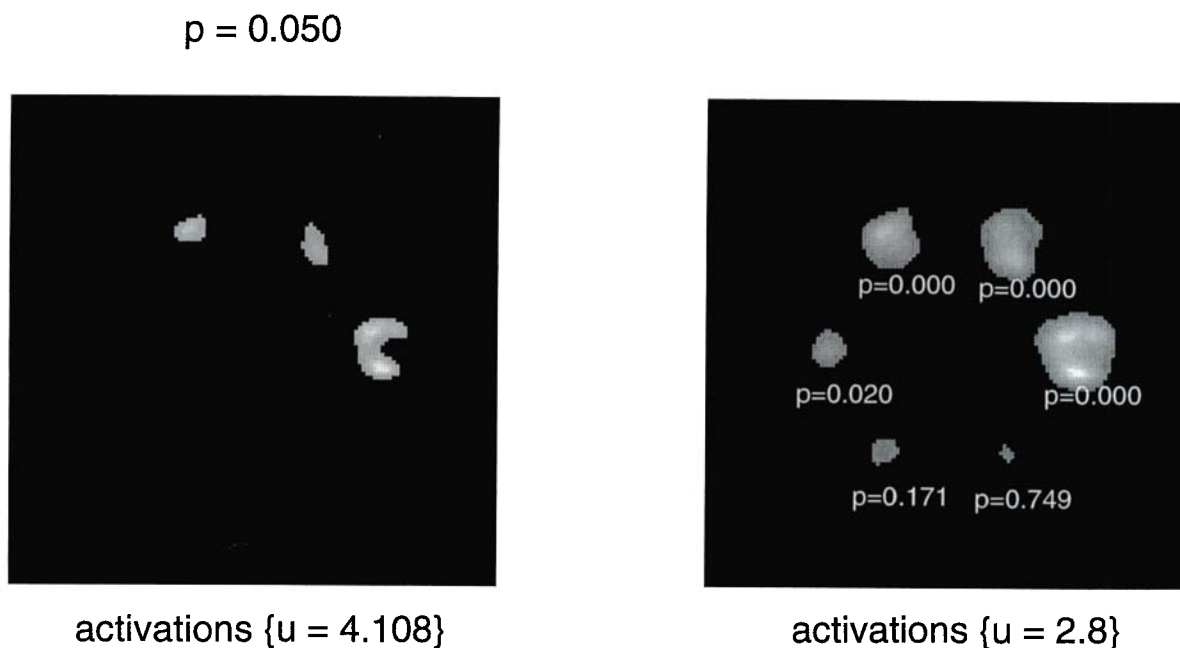
percent higher than in the main body of the phantom. **Bottom left.** Difference image obtained by subtracting the activated image from the baseline image. **Bottom right.** Underlying signal following convolution with the point spread function.

measured with PET is substantially larger ( $\sim 12$  mm) than the autocorrelation of noise ( $\sim 8$  mm) [Friston et al., 1992]. Future work by Siegmund and Worsley will attempt to resolve this “dependency on assumptions,” by searching over tuning parameters (like smoothing) as well as voxels.

The apparent volume of activation in an SPM is not the real volume of activated cerebral tissue. Furthermore, one must also be aware that small but intense physiological activations will be subject to partial volume effects.

We reiterate that the topography of activation foci has been analyzed only with respect to size. Other

features (shape, symmetrical distribution, etc.) may render a particular focus, or set of foci, sufficiently improbable that they can be accepted as real. The scope of analyses that could be brought to bear on statistical parametric maps is clearly extensive. The results presented have, of course, many other potential applications, which will be pursued in subsequent publications. The expressions presented are approximations (usually exact in the limit of high thresholds). It is possible that more exact expressions may become available in the next year or two, or at least other approximate answers should appear to substantiate or supersede the present ones. In the interim, the results



**Figure 8.**

Results of thresholding to detect significant activations. **Left:** Conventional threshold [setting  $P(n_{\max} \geq 1)$  to 0.05] of 4.108. **Right:** A lower and arbitrary threshold (2.8) has been applied to the data, and the significance of each activation focus has been

assessed using  $P(n_{\max} \geq k)$ . This new approach to detecting significant activations is more sensitive because it is not threshold dependent and takes account of the size of each activated region.

presented here appear to perform well enough to justify practical application.

We have deliberately not limited the analysis or simulations to three-dimensional processes because the results are applicable in any dimension. Although we envisage such an approach being applied to three-dimensional SPMs, other applications are important. For example, in assessing peri-stimulus-joint-histograms, obtained from multiunit electrode recording data, the problem is assessing the significance of conjoint unit activity at some temporal distance from stimulus onset. This problem can be formulated in terms of a smooth one-dimensional Gaussian process and can be addressed using the equations presented above. At the other extreme, functional or dynamic MRI studies of the hemodynamic response to a repeating stimulus can be assessed using SPMs, which are also a function of time from stimulus onset (e.g., an SPM of the cross-correlation function between the functional MRI signal and some time-dependent sensorimotor or cognitive parameter). In this instance, an SPM is computed at a series of temporal offsets from the start of stimulation. Because the repeat time of scans is typically less than the hemodynamic time constants, an effective *temporal point spread function*

introduces smoothness in the time domain. A time series of SPMs constitutes a four-dimensional SPM, which could be modeled as a four-dimensional Gaussian process according to the expressions presented here.

#### NOTE ADDED IN PROOF

In this article we presented some theoretical results concerning the spatial extent of activation foci in statistical parametric maps of any dimension. A recent paper [Knorr U, Weder B, Kleinschmidt A, Wirrwar A, Huang Y, Herzog H, Seitz RJ (1993): Identification of task specific rCBF in individual subjects: Validation of an application for PET, JCAT 17:517–528], which examines this issue from an empirical perspective, presented results which speak to a consistency between our theoretical predictions and their experimental findings. Knorr and colleagues formed clusters above a threshold of 30% of the SPM maximum and then fit the empirical distribution of cluster size ( $n$ ) with a gamma distribution. Their results were restricted to two dimensional SPMs. The gamma distribution was a fortuitous choice since this should hold asymptotically in two dimensions (where the distribution should be

exponential). Note that our analysis suggests the gamma distribution is not appropriate for one, three, or more dimensions.

An interesting verification of our results is Fig 3a which is a plot of peak height against mean activity (within each region). If the peak is parabolic, as the theory of Nosko predicts, this should be a straight line with a slope of 2 (irrespective of the FWHM). This appears to fit the data reasonably well. Figure 3b shows a plot of peak height against region size which again should be asymptotically a straight line with a slope of:

$$\text{threshold} \cdot 2 \cdot \ln(2) / (\pi \cdot \text{FWHM}^2)$$

which seems to fit with the FWHM of 5.5 mm.

### ACKNOWLEDGMENTS

Part of this work was performed while K.J.F. was funded by the Wellcome Trust.

### REFERENCES

- Adler RJ, Hasofer AM (1981): *The Geometry of Random Fields*. New York: Wiley.
- Friston KJ, Frith CD, Liddle PF, Dolan RJ, Lammertsma AA, Frackowiak RSJ (1990): The relationship between global and local changes in PET scans. *J Cereb Blood Flow Metab* 10:458–466.
- Friston KJ, Frith CD, Liddle PF, Frackowiak RSJ (1991): Comparing functional (PET) images: The assessment of significant change. *J Cereb Blood Flow Metab* 11:690–699.
- Friston KJ, Frith CD, Passingham RE, Dolan RJ, Liddle PF, Frackowiak RSJ (1992): Entropy and cortical activity: Information theory and PET findings. *Cereb Cortex* 2:259–267.
- Friston KJ, Jezzard P, Frackowiak RSJ, Turner R (1993): Characterizing focal and distributed physiological changes with MRI and PET. In: *Functional MRI of the Brain*. Berkeley: Society of Magnetic Resonance in Medicine, pp 207–216.
- Hasofer AM (1978): Upcrossings of random fields. *Suppl Adv Appl Prob* 10:14–21.
- Nosko VP (1969a): Local structure of Gaussian random fields in the vicinity of high level shines. *Sov Math Doklady* 10:1481–1484.
- Nosko VP (1969b): The characteristics of excursions of Gaussian homogeneous random fields above a high level. In: *Proceedings of the USSR-Japan Symposium on Probability* (Harbarovsk, 1969). Novosibirak, pp 216–222.
- Nosko VP (1970): On shines of Gaussian random fields (in Russian). *Vestn Moscov Univ Ser I. Mat Meh* 1970:18–22.
- Poline JB, Mazoyer BM (1993): Analysis of individual positron emission tomography activation maps by detection of high signal-to-noise-ratio pixel clusters. *J Cereb Blood Flow Metab* 13:425–437.
- Roland PE, Levin B, Kawashima R, Ackerman S (1993): Three dimensional analysis of clustered voxels in <sup>15</sup>O-butanol brain activation images. *Hum Brain Mapping* 1:3–19.
- Worsley KJ, Evans AC, Marrett S, Neelin P (1992): A three-dimensional statistical analysis for rCBF activation studies in human brain. *J Cereb Blood Flow Metab* 12:900–918.
- Worsley KJ (in press): Local maxima and the expected Euler characteristic of excursion sets of <sup>2</sup>, F and t fields. *Adv Appl Prob*.